



Proposal to the U.S. Department of Education
NCLB GROWTH MODEL PILOT PROGRAM

February 16, 2006

Revised March 17, 2006

Revised May 15, 2006

Revised June 1, 2007

EXECUTIVE SUMMARY

INTRODUCTION

On November 21, 2005, Secretary Margaret Spellings requested that states submit proposals to participate in a new NCLB growth model pilot program. In response to this request, Tennessee proposes to use a *projection model* – not a value-added model – to test the efficacy of integrating longitudinal analysis of student achievement data into its NCLB accountability system. This system will encourage schools to put individual students who have yet to reach proficiency on accelerated paths to meeting state achievement standards. It will also encourage schools to identify and provide appropriate interventions to students who are at-risk of falling below proficiency. If approved, the state will implement this system for elementary and middle AYP determinations based on 2005-06 testing.

PROPOSED MODEL

The projection model supplements the statutory AYP model. It uses individual student projection data to determine the percent of students, by subgroup and subject area, who are projected to attain proficiency on the state assessment three years into the future. It uses 7th and 8th grade projections for 4th and 5th grade students, respectively, and uses high school graduation exam projections for 6th – 8th grade students. The model uses current-year scores for 3rd grade students, students new to the state, and students who take alternative assessments.

Schools and districts meet AYP proficiency requirements through the projection model if *all subgroups* meet the annual measurable objective in the content area. Based on analysis of 2004-05 data, the State estimates that approximately 13% (47) of schools that do not meet AYP under the statutory status/safe harbor model will meet AYP with this projection model.

CORE PRINCIPLES

1. The projection model will encourage schools and districts to bring all students to a high standard of proficiency and eliminate gaps in reading/language arts and mathematics.
2. The projection model requires low-achieving students to make accelerated progress toward proficiency and does not alter this expectation based on student characteristics.
3. The proposed accountability system produces separate accountability decisions in reading/language arts and mathematics.
4. The proposed accountability system includes all students in tested grades in the assessment and accountability, holds schools accountable for the performance of student subgroups, and includes all schools and districts.
5. Tennessee has had annual assessments in reading/language arts and math in each of grades 3-8 since 1992, and high school exams since 2001. These assessments produce comparable results from year to year and grade to grade, and are expected to be approved through the peer review process for the 2005-06 school year.
6. The projection model uses individual student projection data derived from the student's prior achievement data. The state's longitudinal data system tracks student progress across time and across schools and districts.
7. The accountability system requires that all subgroups attain a 95% participation rate in each subject area and that all students attain the 93% attendance rate.

I. INTRODUCTION

The No Child Left Behind Act of 2001 (NCLB) launched the United States on a new course to ensure that all students meet a high standard of proficiency in reading/language arts and mathematics by 2013-14. By focusing acute attention on the performance of student “subgroups” – students in poverty, students with disabilities, students with limited English proficiency, and students in racial and ethnic minorities, NCLB has illuminated striking disparities in student achievement across the nation. By compelling schools to make adequate yearly progress (AYP) toward bringing all students to proficiency and prescribing interventions for schools that fall short, NCLB has created incentives and resources to drive schools and engage parents and communities to eliminate the nation’s most fundamental educational inequities.

On April 7, 2005, Secretary Margaret Spellings announced that the U.S. Department of Education would grant states new tools to meet this crucial goal. On November 21, 2005, Secretary Spellings requested that states submit proposals to participate in a pilot program to test the efficacy of incorporating growth models into AYP calculations. Of Tennessee’s two growth models – a value-added model that estimates district, school, and teacher effect scores and a projection model that estimates individual students’ projected scores on future assessments – only one is appropriate for the NCLB growth model pilot program. The value-added model, which measures whether districts, schools, and teachers provide sufficient instruction for their students *as a group* to make *one year of progress* each year, is an innovative mechanism to drive academic progress for all students but is clearly not aligned with NCLB’s precise goal that each individual student will reach proficiency. The projection model, meanwhile, by predicting *each student’s* future achievement *relative to state standards*, holds great promise as a mechanism to guide education policy and practice under NCLB.

In response to the Secretary’s request, Tennessee proposes to use the projection model, rather than the value-added model, to test the efficacy of integrating a growth model into its NCLB accountability system. Tennessee will incorporate individual student projection data into AYP calculations in a manner that supports the “*Bright Lines*” of NCLB and follows the intent of the “safe harbor” exception clause. By incorporating this data into AYP, Tennessee will encourage schools to put individual students who have yet to reach proficiency on accelerated paths to meeting state achievement standards. It will also encourage schools to identify and provide appropriate interventions to students who are at-risk of falling below proficiency. If approved, Tennessee will implement this change for elementary and middle AYP determinations based on testing for the 2005-06 school year.

Policy Rationale for Using a Growth Model in AYP Calculations

Under its current accountability system, Tennessee assigns overall ratings and interventions to schools and districts according to NCLB/AYP statutory requirements. The State also rates schools that fail AYP for the first year as “target” schools, and provides technical assistance to these schools to address the areas where they fell short of AYP standards. The State identifies schools and districts that have missed AYP standards two or more consecutive years in the same content area as “high priority”.

Under the current accountability system, schools meet AYP proficiency standards when all students and subgroups meet annual measurable objectives (AMO’s) in reading/language arts and mathematics proficiency or meet the progress requirements under the “safe harbor” exception clause. The “safe harbor” exception provides that subgroups that have yet to meet the AMO may meet AYP if the subgroup has reduced the percent of students below proficient by 10% from the previous year and made progress on an additional indicator. This accountability system encourages schools and districts to improve student achievement and close achievement gaps by

focusing resources on students in subgroups that have yet to meet annual proficiency targets. While this system has led to substantial educational improvements across Tennessee, it lacks sufficient precision to shape effective and efficient education policy and practice in the years ahead.

By incorporating student projection data into AYP calculations, Tennessee's new accountability system will encourage schools and districts to improve student achievement and close achievement gaps by focusing resources on **all students who have yet to attain proficiency or are at-risk of falling below proficiency**. It will give schools and districts an immediate incentive to identify students who start out far behind and launch them on an accelerated path to proficiency in later grades. It will also compel schools and districts to catch proficient and even advanced students who are slipping over time. The new accountability system will also serve the following purposes:

- Reinforce Tennessee's approach to meeting NCLB goals by assisting educators to differentiate instruction and interventions based on individual student needs. The proposed accountability system is consistent with the State's approach to assisting schools and districts in bringing all students to high standards. The State guides schools and districts to meet these goals by addressing the needs of individual students. It provides intensive professional development and technical assistance to guide educators in using data to identify individual student needs and differentiate instruction based on these needs.
- Press educators, parents, and communities to have high expectations for students who have yet to reach proficiency or are at-risk of falling below proficiency. It will demonstrate that, with appropriate instruction and interventions, individual students **will** make accelerated progress toward meeting state standards.
- Affirm the effectiveness of those "high-impact" schools and districts that provide instruction and interventions to successfully place individual students who have yet to meet proficiency on accelerated paths to meeting state standards.
- Encourage educators to make use of valuable longitudinal assessment data to precisely diagnose and treat individual student needs, and encourage state and local policymakers to use longitudinal assessment data to precisely target interventions and technical assistance.
- Engage parents and communities in the process of using data to provide individual students with the support they need to reach state standards and beyond. The State will work with parent and community groups to educate them about the power of being able to use projections to drive academic improvement and to more precisely measure the impact of schools and districts.
- Target state resources toward districts and schools in the greatest need of assistance to develop and implement effective practices to ensure that all students meet state standards in reading/language arts and mathematics. Tennessee has numerous elementary and middle schools that are making tremendous progress with individual students, and the State strongly prefers to concentrate its resources on assisting other schools in replicating this success.

Tennessee's Actions to Meet NCLB Principles

Under NCLB, Tennessee has taken the following actions to meet the goal of all students reaching proficiency in reading/language arts and mathematics by 2013-14:

- **Implemented numerous initiatives to improve student achievement and close achievement gaps.**

- Trained educators to differentiate instruction based on individual student needs through training sessions for nearly 100% of school districts and personnel from all nine Field Service Centers.
- Partnered with Ruby K. Payne to train teachers, principals and supervisors through a two-day seminar based on Payne's book, *A Framework for Understanding Poverty*. The state also offers an in-depth 'train the trainer' series to equip teachers, principals and supervisors to take what they have learned and implement their own district-level professional development around this framework.
- Established an Urban Education Improvement Office for educators to share resources and ideas on how to address the needs of students in urban areas. More than 1,100 teachers, principals and administrators have attended training, in-service and conference sessions in addition to more than 30 school visits by departmental staff.
- Introduced a national, research-based model for high-quality instruction for English Language Learners.
- Convened a Closing the Achievement Gap Task Force to identify and disseminate best practices for improving performance for special education students.
- Launched the Tennessee Comprehensive System-wide Planning Process (TCSPP), to unify district leaders around common goals to improve student achievement and eliminate achievement gaps.
- Published the *Blueprint for Learning*, a guide to the state curriculum to help teachers know what skills each student should have at each grade level.
- Deployed state assessment personnel to lead Assessment Literacy workshops to train administrators and teachers how to interpret longitudinal student assessment data – including student projections – and use this data to drive district, school, and classroom practice. To date, these workshops have trained 103 of 136 district superintendents, 2249 principals and supervisors, and 791 teachers. The State will hold 10 sessions this summer expected to reach half the teaching force. The State aims to train all teachers by the summer of 2007.

- **Improved student achievement and narrowed achievement gaps.**

Between 2003-04 and 2004-05 in grades 3-8 and in high school, Tennessee saw achievement improve and achievement gaps narrow between white and black students, economically disadvantaged and not disadvantaged students, and students with and without disabilities. Between 2003-04 and 2004-05, the elementary and middle achievement gap in reading/language arts between white and black students closed by 4.9

percentage points. The gap between students based on economically disadvantaged status closed by 5.3 points. The gap between students based on disability status closed by 11.9 points. Over the same time-period, the elementary and middle achievement gap in mathematics between white and black students closed by 4.9 points. The gap between students based on economically disadvantaged status closed by 4.6 points. The gap between students based on disability status closed by 8.4 points.

- **Held schools and districts accountable for the reading/language arts and mathematics performance of all students and subgroups.**

Tennessee has tested all students in grades 3-8 and as they complete the reading/language arts and mathematics graduation exams. The State has applied rules and procedures outlined in the Tennessee Accountability Workbook to this data to determine whether schools and districts, by all students and subgroups, have met annual measurable objectives in reading/language arts and mathematics proficiency. It has also determined whether all students and subgroups have met the 95% participation rate in each subject, and whether schools and districts meet the additional indicator. Using these analyses, the State has then identified schools and districts in need of improvement.

The State has reported this data and other information about NCLB to the public on the “NCLB Reports” website at the end of each summer, and again on the State’s Annual Report Card in late fall. It has provided appropriate interventions and technical assistance to schools and districts identified as in need of improvement. It has reestablished nine Field Service Centers to provide technical assistance to target and high priority schools and placed Exemplary Educators (EE’s), highly-trained veteran educators, in high priority schools.

- **Empowered parents with information and options to improve their children’s educational opportunities.**

- Reporting on district and school academic performance:
 - The State’s Annual Report Card (<http://www.k-12.state.tn.us/rptcrd05/>) includes student assessment data for reading/language arts by district, school, subject, grade, and subgroup. It also includes student assessment data for social studies, science, and writing by district and school, and ACT data by district, school, and subject.
- Reporting on AYP and improvement status:
 - The NCLB Reports Website (<http://www2.state.tn.us/k-12/ayp05.asp>) includes district and school AYP reports, a list of target schools and high priority schools, and background and explanatory information about the state’s accountability system.
 - The Annual Report Card (<http://www.k-12.state.tn.us/rptcrd05/>) also includes each district and school AYP and improvement status, as well as data used to make AYP determinations.
- Public School Choice: Students in all high priority Title I schools are offered school choice. Under Tennessee law, students in non-Title I schools that are in subgroups

that do not meet AYP standards are also eligible for school choice. The State's website includes a list of frequently asked questions that should be helpful to parents and other stakeholders. <http://www.state.tn.us/education/fedprog/fpschlchoice.php>

- Supplemental Education Services: The State's website includes a list of schools required to offer SES and a list of approximately 50 approved providers. <http://www.state.tn.us/education/fedprog/fpses.php>
- **Improved teacher quality and provided information on the quality of local teachers.**
 - The State has offered "highly qualified academies", five-day workshops that have provided in-depth reading/language arts and mathematics training to more than 1100 Tennessee teachers.
 - The State provides information on the number of out-of-field teachers and number of classes taught by highly qualified teachers on its Annual Report Card and information on the educational qualifications of teachers in its Annual Statistical Report.

Core Elements for Growth Models are Met

Tennessee is well-positioned to participate in the growth model pilot program. The Tennessee Comprehensive Assessment Program (TCAP) includes annual testing of students in grades 3-8 using vertically-aligned assessments. The proposed accountability system is supported by a statewide longitudinal student assessment database and a robust statistical methodology. The database includes a unique student identifier that has allowed the State to track students across schools and districts and over time since 1992. The database has also permitted the State to implement a statistical methodology to project individual student scores on future assessments using all of a student's prior achievement data.

In October 2001, the State launched a secure website for principals and teachers to access their students' individual assessment data. In 2002, the State began reporting individual student projections to future achievement levels to help educators identify students in greatest need of assistance to meet or stay at state standards. The projection methodology uses all of an individual student's prior achievement scores to estimate the student's achievement level at a future point in time. The model's only predictor variables are the student's prior test scores. By assuming that the student will have the average Tennessee schooling experience in the future, it includes estimated mean scores for the average school in Tennessee and regression coefficients that are pooled within schools across the state. These coefficients are updated each year as a new student cohort acquires test scores at the projection endpoint. The only source of the model's complexity is missing data – not all students have prior achievement scores for all subjects at all grades/years. Please see the technical appendix for a detailed description of the model and its solution to the missing data problem.

Given Tennessee's historical use of value-added scores for district and school accountability and teacher evaluation, it is important to clarify that neither the proposed projection model for AYP nor its underlying projection methodology relies on a value-added model.

In addition, given the diversity of growth models in use across the nation, it is important to reiterate that Tennessee's projection methodology uses **only** a student's prior test scores as

predictor variables. The student's race/ethnicity, gender, economically disadvantaged status, disability status, and language proficiency status are not included in the model. These student characteristics are not a factor in an individual's expected academic progress and should not be included in such a model.

Other Assurances

- The State has discussed the proposed accountability system with numerous Tennessee educators. It presented the idea of incorporating projections to future achievement levels and received feedback during AYP Workshops in the summer of 2005 and during the state's LEAD conference in fall 2005.
- The State will report aggregate student projection data used to make AYP decisions on its NCLB Reports website and Annual State Report Card. It will clearly label schools and districts that make AYP using projection data. It will collaborate with educators and parents to design these reports.
- The State welcomes the opportunity to collaborate with the U.S. Department of Education in evaluating the effectiveness of the growth model pilot program on Tennessee student achievement.

II. PROPOSED MODEL

Under the proposed accountability system, schools and districts will have two options for meeting elementary and middle AYP proficiency targets in reading/language arts and mathematics:

Status Model

All subgroups meet the annual proficiency targets using the percent of students scoring proficient or advanced based on *current-year, 2-year average, or 3-year average test scores* or meet the requirements of the safe harbor exception clause.

Projection Model

All subgroups meet the annual proficiency targets using the percent of students scoring proficient or advanced based on *projected test scores* three years into the future.

To meet AYP through the projection model, all subgroups must meet the annual measurable objectives using the percent of students projected to score proficient or advanced on the statewide assessment three years into the future. It expects fourth and fifth grade students to make accelerated progress toward attaining proficiency in time to be prepared for high school work. It expects sixth, seventh, and eighth grade students to make accelerated progress toward attaining proficiency on the state's graduation standards. All students' scores will be included in the model. For students in their first tested year in Tennessee, the State will use the student's current-year score. This includes students in the 3rd grade and students who are new to the state. Projected scores that fall above the proficiency standard for the future assessment will be

regarded as proficient (Figure 1). Projected scores that fall below the proficiency standard for the future assessment will be regarded as below proficient (Figure 2).

Figure 1: Gateway Algebra I Report for Student B (Proficient)

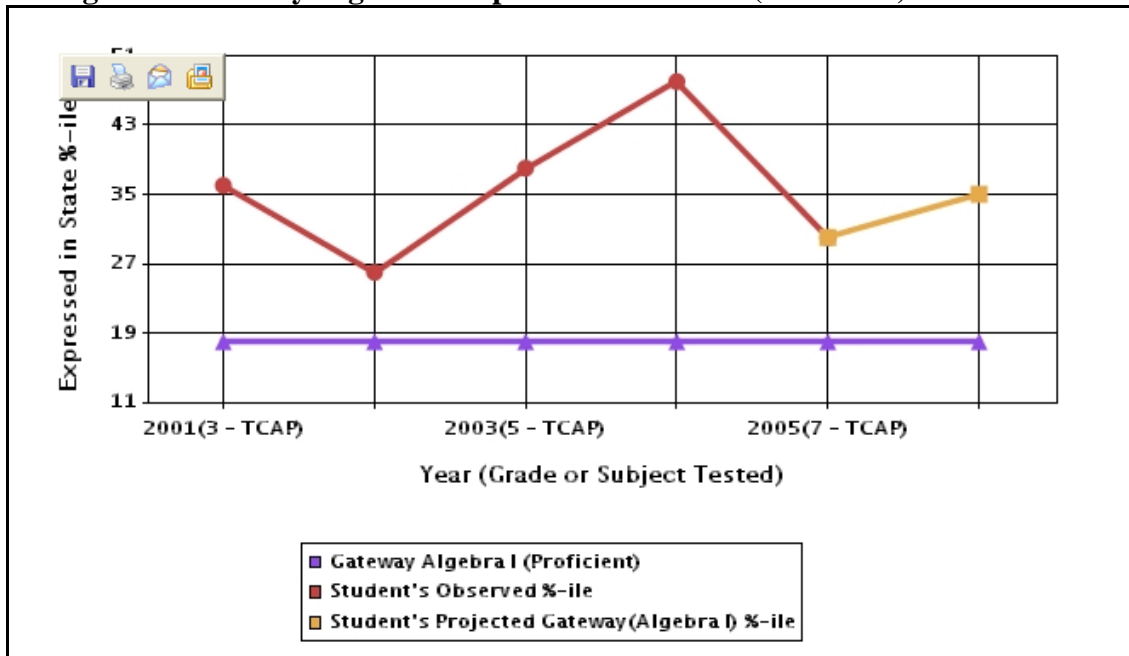
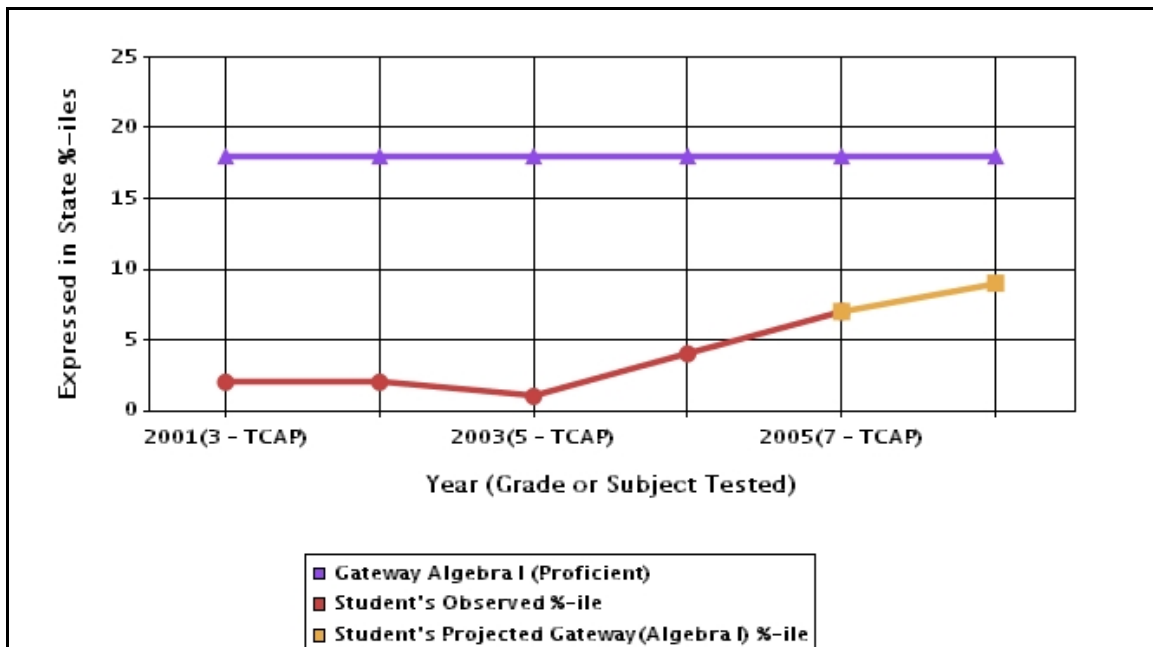


Figure 2: Gateway Algebra I Report for Student A (Below Proficient)



The projection model sets a high standard. Schools and districts may meet AYP proficiency requirements in each content area only if each subgroup's projected percentage of students who score proficient or advanced meets the approved annual measurable objective for that content area. The AMO's have been approved in Tennessee's Accountability Workbook. They increase over time until they reach 100% in 2013-14.

The projection model assigns school credit for all students who are projected to be proficient three years into the future, whether they are currently below proficient or are currently proficient. It does not assign schools **any** credit for students who are currently proficient but are projected to score below proficient on the future assessment. It does not assign schools any additional credit for students who score advanced.

The projection model includes current scores for 3rd grade students and other students who are in their first tested year in Tennessee. If the student scores proficient in the current year, he or she will be counted as proficient in the projection model. If the student scores below proficient in the current year, he or she will be counted as below proficient in the projection model.

The projection model will not apply to high school; however, the State expects that the increased focus on projections will support high school reform efforts to improve high school graduation and college-readiness rates. On February 7, 2006, Tennessee Governor Phil Bredesen called for the state to improve high school graduation rates from 78% to 90%, and college graduation rates from 43% to 55%, by 2012. By examining young students' projections to the high school graduation exams, educators can quickly identify students in greatest need of assistance to meet high school standards. By examining students' projections to ACT, educators can easily identify students who need assistance meeting college-readiness benchmarks.

III. CORE PRINCIPLES

CORE PRINCIPLE 1

The proposed accountability model ensures that all students are proficient by 2013-14 and sets annual goals to ensure that the achievement gap is closing for all students.

1.1 How does the State accountability model hold schools accountable for universal proficiency by 2013-14?

1.1.1 Does the State use growth alone to hold schools accountable for 100% proficiency by 2013-14? If not, does the State propose a sound method of incorporating its growth model into an overall accountability model that gets students to 100% proficiency by 2013-14? What combination of status, safe harbor, and growth is proposed?

The State proposes to use status, safe harbor, and growth to hold schools and districts accountable for 100% proficiency by 2013-14 for elementary/middle grades in reading/language arts and mathematics. The State proposes to use a projection model, rather than a value-added or other form of growth model, to evaluate individual student academic progress toward meeting state standards.

The State will use status and safe harbor to hold schools and districts accountable for 100% proficiency by 2013-14 for high school grades in reading/language arts and mathematics.

1.2 Has the State proposed technically and educationally sound criteria for “growth targets” for schools and subgroups?

1.2.1 What are the State’s growth targets relative to the goal of 100% of students proficient by 2013-14?

The projection model will include all elementary/middle students tested under the Tennessee Comprehensive Assessment Program (TCAP). Table 1 lists the proficiency definition for each student by category. For purposes of the projection model:

- A 4th or 5th grade student will be considered proficient if the student is projected to score above the proficiency standard on the TCAP assessment three years into the future. A 4th or 5th grade student will be considered below proficient if the student is projected to score below the proficiency standard on the TCAP assessment three years into the future. For example, a 4th grade student with a projected 7th grade reading/language arts score that falls above the 7th grade reading/language arts proficiency standard will be counted as proficient. A 5th grade student with a projected 8th grade mathematics score that falls below the 8th grade math proficiency standard will be counted as below proficient in the projection model.
- A 6th, 7th, or 8th grade student will be considered proficient if the student is projected to score above the proficiency standard on the TCAP high school graduation assessment. A 6th, 7th, or 8th grade student will be considered below proficient if the student is projected to score below the proficiency standard on the TCAP high school graduation assessment. For example, a 6th grade student with a projected score on the high school reading/language arts assessment (English II) that falls above the English II proficiency standard will be counted as proficient. A 7th grade student with a projected score on the high school mathematics assessment (Algebra I) that falls below Algebra I proficiency standard will be counted as below proficient in the projection model.
- Students in their first tested year in Tennessee, including 3rd grade students and students with no prior test score, will be considered proficient if they score above the proficiency standard in the current year and considered below proficient if they score below the proficiency standard in the current year.
- Students who take alternative assessments will be considered proficient if they score above the proficiency standard for that alternative assessment and below proficient if they score below the proficiency standard for the alternative assessment. This rule will follow current policy and procedures regarding inclusion of alternate assessment scores in AYP. If these students have taken regular assessments in the past, they may have a projection score; however, these students’ performance

may only be measured appropriately through the alternative assessment and standards.

Table 1: Projection Model Proficiency by Student Category

Student Category	TCAP Score Applied	Proficiency Standard
3 rd grade	3 rd grade	3 rd grade
4 th grade	7 th grade projection	7 th grade
5 th grade	8 th grade projection	8 th grade
6 th -8 th grade	High school projection	High school
With no prior test score	Current score	Current grade
Who take alternative assessments	Current score	Alternative standard

These criteria set a short time-horizon for students to attain proficiency. The model expects 4th and 5th grade students to make accelerated progress towards attaining proficiency by the 7th and 8th grade, respectively. It expects 6th - 8th grade students to make accelerated progress to attain proficiency by the time they take the high school graduation exams for math and reading/language arts, typically during the 9th or 10th grade. This path expects that each student will make substantial progress – much more than a year’s worth of progress – every year until attaining proficiency. By expecting students in greatest need to make the most progress, the proposed model will drive the elimination of student achievement gaps.

1.3 Has the State proposed a technically and educationally sound method of making annual judgments about school performance using growth?

1.3.1 Has the State adequately described how annual accountability determinations will incorporate student growth?

- A. Schools and districts will meet AYP proficiency requirements under the projection model if, *for all subgroups*, the percent of students with proficient scores on reading/language arts meets or exceeds the Annual Measurable Objective (AMO) for the content area. The AMO’s for the projection model are identical to the approved AMO’s in the Tennessee Accountability Workbook (Table 2).

Table 2: Elementary/Middle Annual Measurable Objectives

School Years	Reading/Language Arts Target	Math Target
2005-06 through 2006-07	83%	79%
2007-08 through 2009-10	89%	86%
2010-11 through 2012-13	94%	93%
2013-14	100%	100%

- B. The projection model will use all current rules approved under Tennessee's Accountability Workbook, including disaggregating by subgroup, counting only students with full academic year status, applying a minimum subgroup size of 45 (or 1%, whichever is greater) to assure statistical validity and reliability of AYP decisions based on the projection model.
- C. The State has analyzed elementary and middle school AYP determinations based on 2004-05 test results and the new requirement to include test scores for grades 3-8. It has found that the proposed model would identify approximately 47 additional schools as making AYP (Table 3).

Table 3: School AYP status based on 2004-05 testing, grades 3-8

AYP Model	Overall AYP Status	
	Yes	No
Current	988	353
Proposed	1035	306
<i>Difference</i>	47	-47

1.3.2 Has the State adequately described how it will create a unified AYP judgment considering growth and other measures of school performance at the subgroup, school, district, and state level?

- A. An elementary or middle school will make AYP if it meets all proficiency requirements of the status/safe harbor model or the projection model, meets the 95% participation rate for all subgroups, and meets the additional indicator (attendance rate). A district will make AYP if 1) its elementary/middle level meets all proficiency requirements of the status/safe harbor model or the projection model, meets the 95% participation rate for all subgroups, and meets the additional indicator (attendance rate) *or* 2) its high school level meets all proficiency requirements of the status/safe harbor model, meets the 95% participation rate for all subgroups, and meets the additional indicator (graduation rate).
- B. A subgroup will make AYP if it meets the proficiency requirements of the status/safe harbor model or the projection model and meets the 95% participation rate.
- C. The State will report the results of the status/safe harbor model and the projection model for all elementary/middle schools and districts in a manner that is clear and understandable to the public. These results will be reported on the State's website before the opening of school to provide parents with opportunity to use the information to inform their educational decisions.

1.4 Does the State's proposed growth model include a relationship between consequences and rate of student growth consistent with Section 1116 of ESEA?

- 1.4.1 Has the State clearly described consequences the State/LEA will apply to schools? Do the consequences meaningfully reflect the results of student growth?**

Additional Questions

Please clarify the interventions facing a school or LEA that does not meet AYP under the growth model and whether they are consistent with section 1116?

Schools and districts that do not make AYP for two consecutive years, in the same area (math, reading/language arts, additional indicator), will be identified for improvement and subject to consequences as prescribed in the Tennessee Accountability Workbook. These consequences include parental notification, public school choice, supplementary education services, and other provisions to comply with Section 1116. Schools that do not meet AYP for the first year will be identified as “target” schools and offered State technical assistance.

The projection model will allow the State to focus interventions on schools and districts that need assistance placing individual students on accelerated paths to proficiency and preventing students from falling below proficiency.

By reporting the results of the projection model for all subgroups in elementary/middle schools and districts, the State will also allow the public to recognize schools and districts that are successfully placing individual students on accelerated paths to proficiency and catching students who are at-risk of falling below proficiency.

CORE PRINCIPLE 2

The proposed accountability model establishes high expectations for low-achieving students, while not setting expectations for annual achievement based upon student demographic characteristics or school characteristics.

- 2.1 Has the state proposed a technically and educationally sound method of depicting annual student growth in relation to growth targets?**
- 2.1.1 Has the State adequately described a sound method of determining student growth over time?**
- A. Is the State’s proposed method of measuring student growth valid and reliable?**

The State’s projection model relies on a robust statistical methodology that uses all of an individual student’s prior achievement scores to estimate the student’s achievement level at a future point in time. The methodology has been in use in Tennessee since 2002, when the State began reporting individual student projections on future assessments to inform instructional decisions. The projections to the high school Gateway exams (Algebra I, English II, and Biology) have been of particular importance to educators and students as these exams are required for high school graduation.

The model's only predictor variables are the student's prior test scores. By assuming that the student will have the average Tennessee schooling experience in the future, it includes estimated mean scores for the average school in Tennessee and regression coefficients that are pooled within schools across the state. **These coefficients are updated each year as a new student cohort acquires test scores at the projection endpoint.**

To arrive, for example, at a 6th grade student's projected score on the high school English II exam, the statistical methodology uses scores from students who took the English II exam in the current year who have the same historical pattern of test scores as the 6th grade student. If the student has 3rd grade, 5th grade, and 6th grade scores (but no 4th grade scores), the methodology estimates regression coefficients for these scores based on the subset of students who took the English II exam in the current year who also had 3rd grade, 5th grade, and 6th grade scores (but no 4th grade scores). These coefficients are then applied to the individual student's 3rd, 5th, and 6th grade scores to calculate the student's projected score on the English II exam. If the student has made progress between the 3rd and 6th grade, the model will show if this progress has been sufficient to predict that the student will reach proficiency by the time he or she takes the English II exam.

B. Has the State established sound criteria for growth targets at the student level, and provided an adequate rationale?

The projection methodology's only predictor variables are an individual student's own prior achievement scores. It does not include any student characteristics as predictor variables. In addition, by assuming that the student will have the average Tennessee schooling experience in the future, it includes estimated mean scores based on the average school in the state. The projection methodology sets no expectations based on any student's school, race/ethnicity, gender, poverty status, disability status, or language proficiency status. All students are held to the same high expectations – to achieve proficiency based on the State's achievement standards. The projection model does not assign different values for growth at different achievement levels. The State will continually evaluate the appropriateness of the student growth target criteria, particularly if it makes changes to assessments or content standards.

Additional Questions

1. Could two students with the same reading score in year 1 have different growth expectations in year 2?

The projection model uses all available past scores, not just the previous year. The error of measurement around an individual student's test score

from one administration is often very large. An attempt to measure progress of an individual student to a future meaningful standard will be improved by using the totality of the test data available for each student. By incorporating all the prior test data the covariance structure among test records can be exploited to dampen the error of measurement in any one test score. Given the above, two students with the same set of past scores will have the same projected score since the projected score is determined entirely by the set of past scores. Other information (which school or classroom the student was in, demographic variables, etc.) is not used to make the projections (see also page 7 of the Proposal).

2. **Please clarify the process and procedures for nesting to the school level and explain whether different growth curves will be generated for students from different classrooms or different schools.**

“Nesting down” refers to the use of a “pooled within schools” variance-covariance matrix to produce the “pooled within schools” regression coefficients used for making projections. See Question 5 for additional details. Because NCLB assessments extend only to the school level, not to the classroom level, it is natural to use a school level model for the projections.

The same projection model is used for students from different classrooms or schools as explained in Question 1.

3. **Please clarify the “average schooling experience” noted on page 17 of the proposal and how this will be accounted for in the model.**

As stated on page 27: “Means for an ‘average school’ are obtained by calculating school-mean scores and averaging them over schools.” Professionals within current schools have no direct control over the effectiveness of the schooling that their students will receive when they leave their building and move to other schools. Thus, by developing the models from a pooled within school data structure along with mean scores that are averaged over schools, the projections to future attainment levels for students is based upon the expected attainment level that these students will reach if they have average schooling experiences in the future.

4. **Please clarify what variables will be used to calculate the regression for the growth model.**

See the discussion of errors of measurement in Question 1. For each student, all available past TCAP scores, as far back as grade 3, are used as predictors in the projection model. As explained in Question 5 and in the Technical Appendix, by using a pooled within school variance-covariance structure for all test data from previous cohorts, the projection model regression coefficients that conform to the existing prior data structure for each student can be estimated. By so doing, projections for all students who have prior test data can be made.

5. **Is the proposed model a covariance model, and say more about missing data. Further rationale for school-based averages and whether this is more effective than imputing values.**

As shown on page 26 of the Proposal, the model is somewhat analogous to analysis of covariance in that it combines “regression” with a “grouping” variable (a school effect). For the purpose of making projections into the future, where the school is unknown, the school effect is set to its average value, i.e., zero (“average schooling experience,” see Question 3). Thus no “school effect” appears in the projection equation on page 26 since its value is zero. As in analysis of covariance, the regression coefficients are “pooled within school” regression coefficients.

The missing value problem is handled, as explained on page 26-27 of the Proposal, by computing the “pooled within school” variance-covariance matrix of the predictor and response variables. All variables (Y and Xs) are centered around school means in order to obtain pooled-within-school estimates. The covariance matrix of these centered scores is obtained by maximum likelihood (ML) estimation using the EM algorithm implemented in the MI procedure in SAS/STAT. ML is used because of the pervasiveness of missing data which makes estimation with complete cases only (listwise deletion) or with available cases (pairwise deletion) inadvisable. See R. J. A. Little (1992), Regression with Missing X’s: A Review, *Journal of the American Statistical Association*, vol. 87, pp. 1227-1237; or P. T. von Hippel (2004), Biases in SPSS 12.0 Missing Value Analysis, *The American Statistician*, vol. 58, pp. 160-164. Because the variances and covariances are ML estimates, the resulting regression coefficients are ML estimates, with all their desirable properties. Under the MAR assumption (which is much less stringent than the MCAR assumption), ML estimates are unbiased, and they use all the information available in the data rather than excluding scores of students with incomplete data. Because the ML estimates already use all the information available in the data, there is nothing to be gained by imputation. Imputed values would simply be re-using information that has already been used to obtain the ML estimates.

6. **Additional statistical citations or empirical research that demonstrate where this model has been applied to vertically-equated assessments producing similar results.**

Wright, Sanders, and Rivers (2005, Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards, in R. W. Lissitz (ed.) *Longitudinal and Value Added Modeling of Student Performance*, Maple Grove, MN, JAM Press) conducted simulation studies for the explicit purpose of comparing results from the projection model to results from a more traditional hierarchical liner growth model which, unlike the projection model, (1) requires vertically scaled test scores and (2) requires that an explicit mathematical form be assumed for growth over time (linear growth is commonly assumed). In the first simulation all of the explicit assumptions for a “growth model” were set and the subsequent data were analyzed with both the projection model and the “growth model”. Each

model was *equally effective* in predicting future scores for students when the conditions were set to *favor* the “growth model”.

For the second simulation a slight deviation from the assumed explicit mathematical form was introduced, and again the data were analyzed with both models. The projection model was clearly superior under this circumstance. However, there is a case in which the “growth model” would be superior to the projection model. To obtain the coefficients for the projection model, the data from the most recent cohort of students are used. If for some reason the scales are not consistent between adjacent cohorts, then the parameters for the projections could be affected. However, this is of lesser concern because all of the AYP measures are predicated on providing consistent scales across years for each grade and subject. Tennessee like other states will be monitoring for stability of scales to insure that measures of proficiency have the same interpretability across cohorts. Considering the simulation results and Tennessee’s experience with projections, it is felt that the projection model approach is more robust.

CORE PRINCIPLE 3

The proposed accountability model produces separate accountability decisions about student achievement in reading/language arts and in mathematics.

3.1. Has the State proposed a technically and educationally sound method of holding schools accountable for student growth separately in reading/language arts and mathematics?

3.1.1. Are there any considerations in addition to the evidence presented for Core Principle 1?

Under the projection model, the State will apply projected scores in the same content area. To determine whether a school/district/subgroup met the annual proficiency target in reading/language arts, it will use student projected scores in reading/language arts. To determine whether a school/district/subgroup met the annual proficiency target in mathematics, it will use student projected scores in mathematics.

The State’s projection methodology is very flexible. It does not require vertically-linked data nor does it assume a specific growth function (see Technical Appendix for the model). In order to increase reliability and dampen measurement error, the projection methodology uses all of a student’s prior achievement scores from all assessments to project future scores. Given that prior scores from assessments in the same content area have the greatest predictive power, projected scores are largely determined by a student’s prior achievement in the same content area.

In small schools and schools with high mobility, projected scores are more valid measures of school performance than current-year scores because they incorporate all of a student’s prior achievement data. The State’s longitudinal database follows students across time and across Tennessee, maximizing the reliability of the projections for these schools.

CORE PRINCIPLE 4

The proposed accountability model ensures that all students in the tested grades are included in the assessment and accountability system. Schools and districts will be held accountable for the performance of student subgroups. The accountability model, applied statewide, will include all schools and districts.

4.1. Does the State’s growth model address the inclusion of all students appropriately?

4.1.1. Does the State’s growth model address the inclusion of all students appropriately?

- The State does not impute missing data in the projection methodology. Tennessee’s projection methodology includes specialized treatment to solve the missing data problem, allowing it to exploit all of a student’s prior achievement data, even when the student does not have a “full record” of test scores in every subject in every grade/year (see Technical Appendix for model). Tennessee’s longitudinal database dampens missing data problems due to student mobility because it tracks students across time and across the state.
- The State will include current year scores of students assessed under alternate standards where these scores are permitted to be utilized in AYP decisions under current policy.
- The State’s definition of Full Academic Year is continuous enrollment in the school/district since the 1st reporting period. This definition does not need to be modified for the projection model.
- The State will include current-year scores of 3rd grade students and students new to the State.
- The projection model will include projected scores for students who are promoted at mid-year, just as it includes projected scores for students who are missing an assessment.

Additional Question

Please clarify whether the growth model will be applied to all students in every school in the state.

All elementary/middle students will be included in the State’s projection model. If students do not have a projected score, the model will use their current score. Please see Principle 1.2.1.

4.1.2. Does the State’s growth model address the inclusion of all subgroups appropriately?

- The projection model holds schools accountable for the achievement of all subgroups. All subgroups must meet the AMO in the content area for that year.
- Student scores, whether current or projected, will be assigned to the subgroup to which the student belongs in the current year.
- In 2005-06, the State plans to include a separate subgroup for students displaced by Hurricanes Katrina and Rita in accordance with the Secretary's guidance of September 29, 2005. These students will be included only in this subgroup, and this subgroup will not be used for making AYP determinations. The projection model will not include students in this subgroup. However, the State has taken particular care to include these students in the state's assessment system. Each student that has been displaced by the hurricanes will be coded with the required demographic information so that the State may track the subgroup.

Additional Question

Please clarify whether the proposal includes only the current year of data from the alternate assessment. Are additional years of data on the alternate assessment available to be included?

The projection model will include current year scores from students who participate in the alternate assessment. If these students have taken regular assessments in the past, they may have a projection score; however, these students' performance may only be measured appropriately through the alternative assessment and standards.

4.1.3. Does the State's growth model address the inclusion of all schools appropriately?

- All schools and districts receive an AYP determination each year, with the exception of new schools. The State tracks accountability with students rather than school number, so if a school receives a new school number and/or name but serves a preponderance of the same students, the State does not consider it a new school and continues to follow its accountability. For example, if a school in School Improvement 2 gets a new number and name, but serves the same students, it will receive an AYP determination and it can move to Correction Action.
- The State holds K-2 schools accountable based on their receiving school's AYP determination and improvement status. Schools with a single tested grade are held accountable based on that grade's performance. Schools with a single non-tested grade are held accountable based on their receiving school's AYP determination and improvement status.
- Under the projection model, each student has his or her own projected score, so the state will apply that score to the school the student currently

attends. Boundary changes, grade reconfigurations, school closings, and new schools will not preclude a projection for schools.

CORE PRINCIPLE 5

Annual assessments in reading/language arts and math in each of grades 3-8 and high school must have been administered for more than one year, must produce comparable results from year to year and grade to grade, and must be approved through the peer review process for the 2005-06 school year.

5.1. Has the State designed and implemented a Statewide assessment system that measures all students annually in grades 3-8 and one high school grade in reading/language arts and mathematics in accordance with NCLB requirements for 2005-06, and have the annual assessments been in place since the 2004-05 school year?

5.1.1. Provide a summary description of the Statewide assessment system with regard to the above criteria.

In 1990, the Tennessee Comprehensive Assessment Program (TCAP) began annual testing of students in grades 2- 8 in mathematics, reading, language, social studies, and science. Since 2001-02, TCAP has tested grades 3-8 in reading/language arts, mathematics, science, and social studies; grades 5, 8, and 11 in writing; high school Algebra I, English II, and Biology I (Gateway exit exams); and high school Math Foundations II, English I, Physical Science, and U.S. History. The State produces district, school, and individual student reports for each of these assessments.

5.1.2. Has the State submitted its Statewide assessment system for NCLB Peer Review and, if so, was it approved for 2005-06?

The State submitted evidence of its compliance with NCLB standards and assessment requirements in January 2006. It expects to learn the results no later than May 2006.

5.2. How will the State report individual student growth to parents?

The State reports longitudinally-linked individual student achievement data, including projections to future assessments, to each student's district, school, and teachers via a secure website and makes a printable version available for distribution to parents. The projections show each student's predicted score on all future state assessments, by subject and grade, in comparison to the state's standards for proficient or advanced. The projections also show each student's predicted score on the ACT assessment, by subject and composite, in comparison to ACT college-readiness benchmarks. The State provides intensive training to educators to assist them in using this data to improve instruction and identify students in need of extra assistance to meet state standards. It also encourages schools to share this data with parents and students through printable reports.

5.3. Does the Statewide assessment system produce comparable information on each student as he/she moves from one grade level to the next?

5.3.1. Does the State provide evidence that the achievement score scales have been equated appropriately to represent growth accurately between grades 3-8 and high school?

Please see Technical Appendix.

5.3.2. If the State uses a variety of end-of-course tests to count as the high school level NCLB test, how would the State ensure that comparable results are obtained across tests?

N/A

5.3.3. How has the State determined that the cut-scores that define the various achievement levels have been aligned across the grade levels? What procedures were used and what were the results?

Please see Technical Appendix.

5.3.4. Has the State used any “smoothing techniques” to make the achievement levels comparable and, if so, what were the procedures?

Smoothing techniques are not used.

5.4. Is the Statewide assessment system stable in its design?

5.4.1. To what extent has the Statewide assessment system been stable in its overall design during at least the 2004-05 and 2005-06 academic terms with regard to grades assessed, content assessed, assessment instruments, and scoring procedures?

The Tennessee Education Improvement Act of 1992 mandated the administration of assessments to grades 3-8 in mathematics, reading/language arts, science, and social studies as well as specified high school subject areas. In the *High School End of Course Tests Policy*, renamed the *High School Examinations Policy* in August 2002, the State Board stipulated that beginning with students entering the 9th grade in 2001-2002, students must successfully pass examinations in three subject areas - Mathematics, Science, and Language Arts - in order to earn a high school diploma. These examinations, called Gateway Tests, were intended to raise the academic bar for all high school students and add accountability for students' academic performance. In the 2001-2002 school year, the Department of Education began to administer the Gateway Tests three times annually to accommodate students completing work in the fall, spring, and summer semesters.

Both the 3-8 as well as the high school assessments are criterion referenced (CRT, selected response) aligned to the state content standards. Test specifications require content coverage at the state performance indicator (spi) level. An external alignment study completed by Norman Webb in December 2005 documented alignment criteria were met by Tennessee's 3-8 and High School Gateway assessments used to underpin AYP calculations in math and reading/language arts.

Tests are physically scanned and scan files edited at the Tennessee Test Processing Center. A high level of QA is maintained during every phase of this operation. Clean files are then exported to the vendor for application of the scoring algorithms. Data questions and cleaning can be easily accomplished via communication with the vendor by the state editors involved in the scan file creation. Please find outlined in a previous part of this section the standard psychometric protocol used for scale score determination. These assessments are selected response instruments which eliminate concerns associated with inter-rater reliabilities due to training or other potential sources of human error. The scoring procedures and scale score determination protocol have not changed during this test series.

5.4.2. What changes in the Statewide assessment system's overall design does the State anticipate for the next two academic years with regard to grades assessed, content assessed, assessment instruments, scoring procedures, and achievement level cut-scores?

The State does not anticipate any changes to the assessment system's overall design in the next two academic years.

CORE PRINCIPLE 6

The accountability model and state data system must track student progress.

6.1. Has the State designed and implemented a technically and educationally sound system for accurately matching student data from one year to the next?

6.1.1. Does the State utilize a student identification number system or does it use an alternative method for matching student assessment information across two or more years?

The State uses a multi-element student merge key consisting of a unique numeric student identifier, first name, last name, middle initial, birth data, gender and ethnicity codes.

6.1.2. Is the system proposed by the State capable of keeping track of students as they move between schools or school districts over time? What evidence will the State provide to ensure that match rates are sufficiently high and also not significantly different by subgroup?

Tennessee has successfully followed the academic progress of students across all districts within the state since 1992. The State has been merging, storing, retrieving, and analyzing longitudinal student data since 1992 to produce district, school, and teacher effect scores in compliance with TCA 49-1-603 through TCA 49-1-608. It has been analyzing this longitudinal student data since 2002 to produce individual student projections to future achievement levels. It has also been reporting student-level data to educators on a restricted website since 2001.

6.1.3. What quality assurance procedures are used to maintain accuracy of the student matching system?

To further ensure the quality of the data linking, the State applies other algorithms, such as same Soundex codes for name spellings, similar numeric id's (truncated digits or id's with most digits consistent), and reasonableness of cohort membership. Pre-slugged answer documents are increasingly used in Tennessee, and this has dramatically improved the quality of the data available for merging.

6.1.4. What studies have been conducted to demonstrate the percentage of students who can be matched between two academic years? Three years or more years?

In 2005, the merge rate for grades 3-8 with student records of three prior years was 92.3%. An example of the data cleaning that takes place prior to the merge is the problem of duplicate numeric id's for 2 sets of test scores. During the 2005 merge, about 1800 students, approximately 1% of the students tested, had numeric identifiers determined to be invalid because identical identifiers were attached to 2 students' records. In these instances, the numeric identifiers were ignored, and the other elements of the merge key were used to successfully merge the 2005 data with that of previous years. The 2005 data quality improved slightly when compared to that delivered for the 2004 processing. In 2004, approximately 2% of the student records were affected by duplicate numeric ids.

When Tennessee began online reporting of student scores longitudinally linked and reported with the most current demographic information available in each student's test record, the State's merging procedures passed the ultimate test of reasonableness—the scrutiny of the teachers who taught the students. Since the student level reporting began, educators have not reported errors in merging that would have linked one child's record to that of a second child.

Additional Question

Please provide additional information on the match rates for two and three years for the whole population and by subgroup.

		2005 Merge Rates	
		2 academic years	3 academic years
Total Population		95.2	92.3
American Indian/Alaska Native	0.2	92.5	91.4

Asian/Pacific Islander	1.3	91.8	90.1
Black, not Hispanic	24.8	96.8	95.7
Hispanic	3.6	93.7	90.9
White, not Hispanic	69.9	95.0	94.4
Limited English Proficient	2.2	89.1	85.2
Students with Disabilities	15.9	95.3	94.7
Economically Disadvantaged	52.1	95.7	94.8

6.1.5.Does the State student data system include information indicating demographic characteristics (e.g., ethnic/race category), disability status, and socio-economic status (e.g., participation in free/reduced price lunch)?

Yes. It is used for reporting.

6.1.6.How does the proposed State growth model adjust for student data that are missing because of the inability to match a student across time or because a student moves out of a school, district, or the State before completing the testing sequence?

The State’s statistical methodology estimates projection scores for all students who have prior years of data, even if students have missing records. If a student does not have any prior data, the projection model will use the student’s current-year score.

CORE PRINCIPLE 7

The accountability system must include student participation rates in the state's assessment system and student achievement on an additional academic indicator.

The projection model only applies to reading/language arts and mathematics proficiency. Schools and districts with subgroups that do not meet the 95% participation rate or the 93% attendance rate requirements will not make AYP.

IV. ADDITIONAL QUESTIONS

1. The status model will continue to use uniform averaging across two and three years. The projection model will not use uniform averaging.
2. The minimum group size will continue to be 45 (or 1%, whichever is greater) and the projection model will apply this policy.
3. The confidence interval will continue to be 95% but the projection model **will not** apply this policy.
4. The projection model will use projected scores for students who took a regular assessment in the current year. It will use current-year alternative assessment scores for students who took these exams, should these scores’ inclusion fall under current policy.

5. The projection model includes projected scores of all students (subject to the exemptions described above). Students whose score is above the cut for proficiency will be counted as proficient. Students whose score is below the cut for proficiency will be counted as below proficient. It does not “credit” schools for students who have projections above proficiency.
6. The State will publicly report data from the projection model in a manner consistent with its traditional reporting of AYP data, substituting aggregate projection scores. It will continue to make individual student projection data available to educators to use in instruction and to share with students and parents. It looks forward to participating in the U.S. Department of Education’s evaluation initiatives.

V. CONCLUSION

Tennessee’s proposed model reflects the “*Bright Lines*” of NCLB, encouraging elementary and middle schools to improve student achievement and close achievement gaps by targeting effective instruction and services to students in greatest need. Schools that are successfully implementing these practices are placing and keeping all students on individual, accelerated paths to attaining high academic standards. The proposed model will validate community and parent perceptions that these are effective schools, and will allow the State to focus interventions on schools that need the most assistance in replicating these effective practices. It will allow Tennessee educators to complete their extraordinary work in narrowing achievement gaps; to finally, by 2013-14, ensure that all students are performing at high standards.

TECHNICAL APPENDIX

I. Projection Methodology

From Wright, Sanders, and Rivers (2005, “Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards”, in R. W. Lissitz (ed.) *Longitudinal and Value Added Modeling of Student Performance*, Maple Grove, MN, JAM Press).

The projection methodology estimates an individual student’s academic achievement level at some point in the future under the assumption that this student will have an average schooling experience in the future. The basic methodology is simply to use a student’s past scores to predict (“project”) some future score. At first glance, the model used to obtain the projections appears to be no more complex than “ordinary multiple regression,” the basic formula being:

$$\text{Projected_Score} = M_Y + b_1(X_1 - M_1) + b_2(X_2 - M_2) + \dots = M_Y + \mathbf{x}_1^T \mathbf{b}$$

where M_Y , M_1 , etc. are estimated mean scores for the response variable (Y) and the predictor variables (Xs). However, several circumstances cause this to be other than a straightforward regression problem.

1. Not every student will have the same set of predictors; that is, there is a substantial amount of “missing data.”
2. The data are hierarchical: students are nested within classrooms, schools, and districts, and the regression coefficients need to be calculated in such a way as to properly reflect this.
3. The mean scores that are substituted into the regression equation also must be chosen to reflect the interpretation that will be given to the projections.

As noted above, a projection is the score that a student would be expected to make assuming that the student has the average schooling experience in the future. The means should therefore be those of an average school within the population of schools of interest. Also, given this interpretation, the nesting needs to be carried only to the school level (students within schools); it is not necessary to carry it to the classroom level.

The missing data problem can be solved by finding the covariance matrix of all the predictors plus the response, call it \mathbf{C} , with submatrices \mathbf{C}_{XX} , \mathbf{C}_{XY} (and $\mathbf{C}_{YX} = \mathbf{C}_{XY}^T$), and \mathbf{C}_{YY} . The regression coefficients (slopes) can then be obtained as $\mathbf{b} = \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY}$. For any given student, one can use the subset of \mathbf{C} corresponding to that student’s set of scores to obtain the regression coefficients for projecting that student’s Y value. Because of the

hierarchical nature of the data (the second problem), the covariance matrix \mathbf{C} must be a pooled-within-school covariance matrix. We obtain this matrix by maximum likelihood estimation using an EM algorithm (to handle missing values) applied to school-mean-centered data. Means for an “average school” are obtained by calculating school-mean scores and averaging them over schools. For brevity, we refer to the elements of \mathbf{C} , along with the vector of estimated means, as the “projection parameters.” Generally, we obtain the projection parameters using the most recent year’s data. That is, we use students who have a Y value in the most recent year and X values from earlier years to get the projection parameters. Projections are then obtained by applying these parameters to students who have X values in the current year (and earlier years) but no Y value.

This methodology does not require vertically linked data nor does it need to assume a linear growth function (or any other specific growth function). Instead, what is required are good predictors of the response variable. The predictors need not be on the same scale with the response or with one another. Potentially, they could be test scores from different vendors and even in different subjects from the response. This gives the methodology considerable flexibility.

II. Comparable Results (5.3.1)

Tennessee criterion referenced assessments for grades 3-8 provide student performance data based upon vertical scales that were developed utilizing industry standard procedures. Equivalent scales are developed for each subsequent operational test form.

As in the TCAP-O CRT assessment, the items in the TCAP-P CRT assessment are all selected-response items. To analyze these items, the three-parameter logistic (3PL) model (Birnbaum, 1968; Lord, 1980) was used. In the 3PL model, the probability that an examinee with scale score θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response by a low-scoring examinee.

Parameter estimations of the 3PL model (and other IRT models) were implemented using CTB’s PARDUX software (Burket, 1991). PARDUX estimates parameters simultaneously for dichotomous and polytomous items using marginal maximum likelihood procedures implemented with the EM algorithm (Bock & Aitkin, 1981; Thissen, 1982). PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. Extensive simulation studies and comparisons between PARDUX and MULTILOG (Thissen, 1990), a program widely

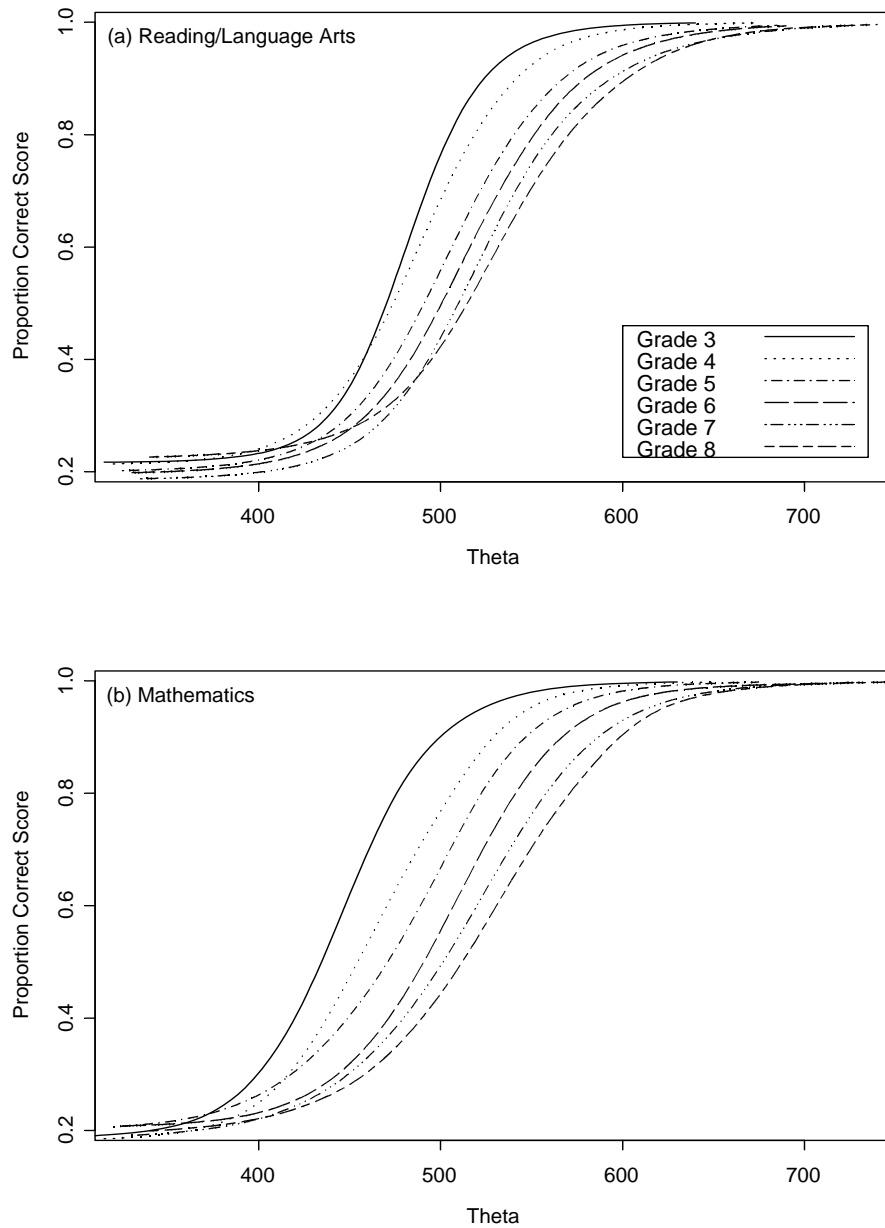
used for research purposes, have shown that PARDUX provides precise estimates of the item and ability parameters, and it performs more efficiently than MULTILOG (Fitzpatrick, 1991). Simulation studies have also compared PARDUX with PARSCALE (Muraki & Bock, 1991), and with BIGSTEPS (Wright & Linacre, 1992). Fitzpatrick and Julian (1996) found that PARDUX provided precise item and ability parameter estimates, and performed more efficiently than the other programs. Extensive studies involving simulated data have also shown that the IRT vertical scaling procedures as implemented in PARDUX produce accurate results (Yen & Burket, 1997). The Stocking and Lord (S&L) procedure (Lord, 1983) was used to place the estimated parameters on the scale from which the anchor items (i.e., CAT/5) were drawn.

Custom Vertical Scale for Mathematics and Reading/Language Arts

The custom vertical scales for Mathematics and Reading/Language Arts were established in 2004 for TCAP-O operational items using a Common Linking Blocks Design. The embedded field test items in 2004 were placed in the same vertical scale as the operational items using the equating procedure of Stocking and Lord (Lord, 1983) and using the software Pardux (Burket, 1991). The equating was done by first calibrating all of the TCAP-O items, operational and field test items, combined. Then these items were equated using the operational items, which are already vertically scaled, as anchor items. The equated field test items together with the operational items served as the item pool for selecting the 2005 operational items. Figure 1 shows the test characteristic curves across all grades for Reading/Language Arts and Mathematics. As expected, the curves are of the same shape and are spaced progressively across the grades as a result of the vertical equating.

Figure 1.
Test Characteristic Curves for Reading/Language Arts and

Mathematics



In Spring 2001 (Mathematics and Science) and Spring 2002 (Language Arts), pilot test forms were administered to Tennessee students and calibrated for each content area using a common item equating design. Instead of equating forms sequentially, all forms were calibrated concurrently using all anchor items. Five calibration forms were selected for operational assessments and have been used sequentially in operational assessments starting in Fall 2001 for Mathematics and Science and Fall 2002 for Language Arts¹.

Although the forms have been pre-equated using the calibration data, the anchor items used to link each pair of adjacent forms remained in the operational forms. The anchor items can be used to perform post-equating of operational forms using data obtained from operational assessments.

The *Gateway* high school assessments were scaled and calibrated using item response theory (IRT) procedures and the three-parameter logistic model (Birnbbaum, 1968; Lord, 1980). The three-parameter logistic model (3PL) defines performance on a selected-response item in terms of three item parameters: item difficulty or location, item discrimination, and level of guessing. Introductory discussions of IRT can be found in measurement literature such as *Educational Measurement* (Linn, 1989), or *Introduction to Measurement Theory* (Allen & Yen, 1979; Chapter 11). In the three-parameter logistic model (Birnbbaum, 1968; Lord, 1980), the probability that a student with proficiency θ will responded correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]}$$

where a_i denotes the item discrimination, b_i the item difficulty, and c_i the pseudo-guessing factor or probability of a correct response by a very low-scoring student.

Item Calibration

Gateway tests were administered three times in 2004-2005 academic year-Fall 2004, Spring 2005, and Summer 2005. Each test contains 62 selected-response items including 55 pre-equated, operational items and seven field test items. The 55 pre-equated items had been field tested either through calibration tests in 2001 Spring or through the use of embedded field test items in the operational test between fall 2001 and spring 2004. The items included in the calibration test were calibrated in concurrent calibration design using common items in all six calibration forms for each subject. The items field tested through 2001-2004 operational tests were calibrated using the 55 operational items as anchor. The highest obtainable scale scores (HOSS) and lowest obtainable scale scores (LOSS) were set for each scale.

For operational items appearing on the 2004 and 2005 *Gateway* forms, the IRT models were implemented using PARDUX software (Burket, 1991). PARDUX estimates parameters simultaneously for dichotomous items using marginal maximum likelihood (MML) procedures implemented with the EM algorithm (Bock & Aitkin, 1981; Thissen, 1982).

The Division of Assessment, Evaluation, and Research also conducts extensive equating studies annually with a statistically appropriate sample of assessment data from school systems.

III. Achievement Score Scale and Cut-Score Equating (5.3.3)

Scale Score Estimation

A variety of item response theory (IRT) scoring procedures are available for estimating examinee trait values. The maximum likelihood estimation (MLE) procedure known as “item-pattern” (IP) scoring finds a unique maximum likelihood (ML) scale score estimate for each pattern of scored (e.g., right or wrong) item responses. Estimates based on a sum of item responses, or “number correct scoring” (NC) scoring finds a ML scale score estimate for each number-correct score. The

two procedures are based on the same IRT model and item parameter estimates (e.g., difficulty, discrimination, and guessing). NC scale scores have been found to be *tau*-equivalent to IP scale scores (Yen, 1984); that is, examinees expect to receive the same score on the average from the two procedures.

The NC scoring procedure considers the number of items an examinee answered correctly in determining his/her trait score (θ). The likelihood of a summed score can be obtained as the sum of the likelihoods of all the response patterns that have the same summed score:

$$L_x(\theta) = \sum_{X=x} L(u | \theta)$$

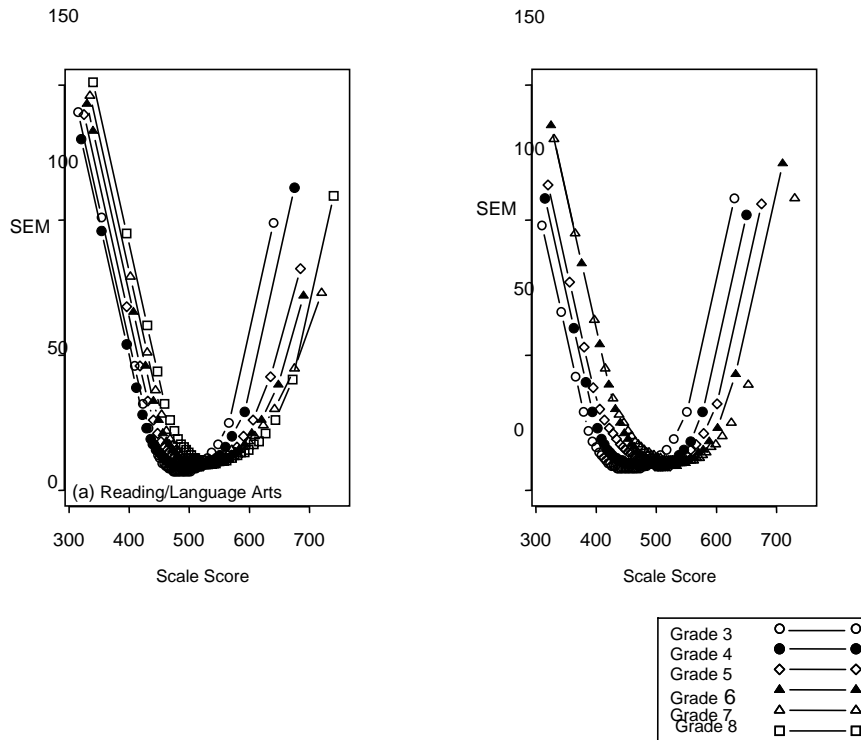
$$X = \sum_{i=1}^n u_i$$

where u_i is a score in item i , and $L_x(\theta)$ is the likelihood function of X , i.e., all possible response patterns that yield a summed score of X . Lord and Wingersky (1984), Hanson (1994), and Thissen and Orlando (2001) described a simple recursive algorithm for the computation of the likelihood function of summed scores.

As in 2004, NC scoring was employed for the 2005 TCAP-P 3-8 CRT assessments in order to accommodate the decision made by Tennessee to report both number-correct and scale scores in individual student reports, and to simplify the scoring process. Number-Correct to Scale Score with SEM Tables reveal that the assessments are measuring very well.

Figure 2 graphically displays standard error of measurement curves for each 2005 TCAP-P CRT assessments. The curves revealed that the tests are measuring very well at the cut-scores, e.g., there is a small standard error of measurements around the cut-scores.

Figure 2
IRT Standard Error Curve



The scale score cuts established by the Tennessee Department of Education (TDOE) with input from the Technical Advisory Committee (TAC) for the High School *Gateway* tests were as follows:

- Mathematics: Proficient = 494, Advanced = 539
- Language Arts: Proficient = 454, Advanced = 511

For each subject, IRT equating procedures have been used to ensure the scale scores are equivalent across forms. Thus, while the raw scores corresponding to the above-described scale score cut-point vary over forms, these raw score cut-points refer to equivalent ability levels. Table 1 displays the score ranges for the three performance levels in scale score and raw score units for each form for each subject. Note: when a scale score cutpoint falls between entries in a Number Correct-to-Scale Score table, the number-correct score with an associated scale score that is closest to the scale score cut-point is used as the performance criterion.

Table 1
Performance Standard for High School *Gateway* 2001-2005

Content Area	Administration	Form	Scale Score			Raw Score		
			Below Proficient	Proficient	Advanced	Below Proficient	Proficient	Advanced
Mathematics	Fall 2001	A	Below 494	494-538	539+	0-29	30-40	41-55
	Spring 2002	B	Below 494	494-538	539+	0-30	31-40	41-55
	Summer 2002	C	Below 494	494-538	539+	0-31	32-41	42-55
	Fall 2002	D	Below 494	494-538	539+	0-29	30-41	42-55
	Spring 2003	E	Below 494	494-538	539+	0-30	31-40	41-55
	Summer 2003	F	Below 494	494-538	539+	0-29	30-41	42-55
	Fall 2003	G	Below 494	494-538	539+	0-29	30-40	41-55
	Spring 2004	H	Below 494	494-538	539+	0-29	30-41	42-55
	Fall 2004	J	Below 494	494-538	539+	0-29	30-41	42-55
	Spring 2005	K	Below 494	494-538	539+	0-29	30-41	42-55
	Summer 2005	L	Below 494	494-538	539+	0-29	30-41	42-55
	Braille	Z	Below 494	494-538	539+	0-29	30-40	41-55
Language Arts	Fall 2002	A	Below 454	454-510	511+	0-25	26-38	39-55
	Spring 2003	B	Below 454	454-510	511+	0-27	28-40	41-55
	Summer 2003	C	Below 454	454-510	511+	0-27	28-40	41-55
	Fall 2003	D	Below 454	454-510	511+	0-24	25-38	39-55
	Spring 2004	E	Below 454	454-510	511+	0-26	27-40	41-55
	Fall 2004	G	Below 454	454-510	511+	0-25	26-39	40-55
	Spring 2005	H	Below 454	454-510	511+	0-24	25-38	39-55
	Summer 2005	I	Below 454	454-510	511+	0-23	24-37	38-55
	Braille	Z	Below 454	454-510	511+	0-25	26-38	39-55